

## G06\_HW4

**Shashank Shekhar (sshekha4)**

**Rahul Aettapu(raettap)**

**Purva Vasudeo (ppvasude)**

- Q1) Starting Cluster Centroids:  
C-1 Centroid: H  
C-2 Centroid: C  
C-3 Centroid: I

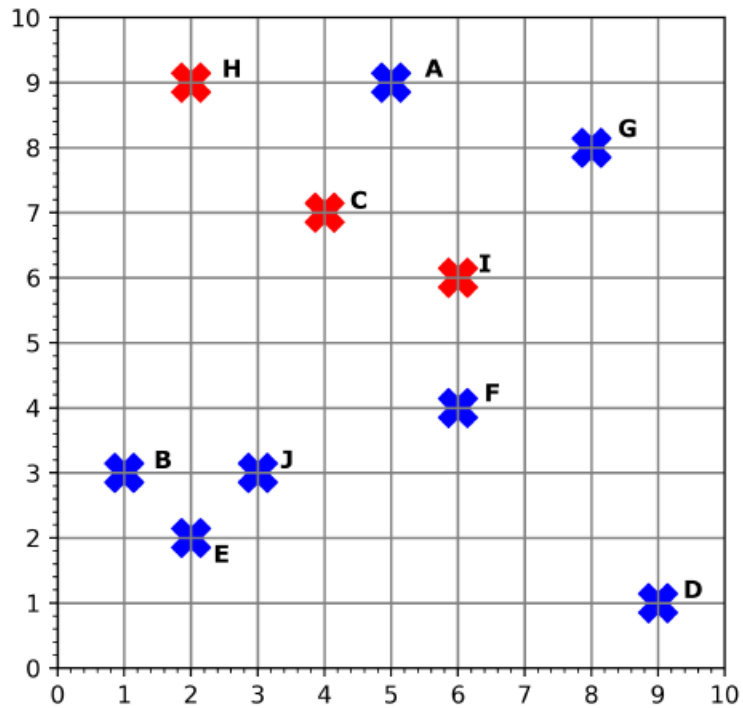


Figure 1: K-means Clustering (a)

### Round 1:

	A	B	C	D	E	F	G	H	I	J
H (2,9)	3.00	6.08	2.83	10.63	7.00	6.40	6.08	0	5.00	6.08
C (4,7)	2.24	5.00	0	7.81	5.39	3.61	4.12	2.83	2.24	4.12
I (6,6)	3.16	5.83	2.24	5.83	5.66	2.00	2.83	5.00	0	4.24

#### Cluster Members:

C-1 members: {H}

C-2 members: {A, B, C, E, J}

C-3 members: {D, F, G, I}

#### New Centroids:

C-1 Centroid: (2,9)

C-2 Centroid: (3, 4.8)

C-3 Centroid: (7.25, 4.75)

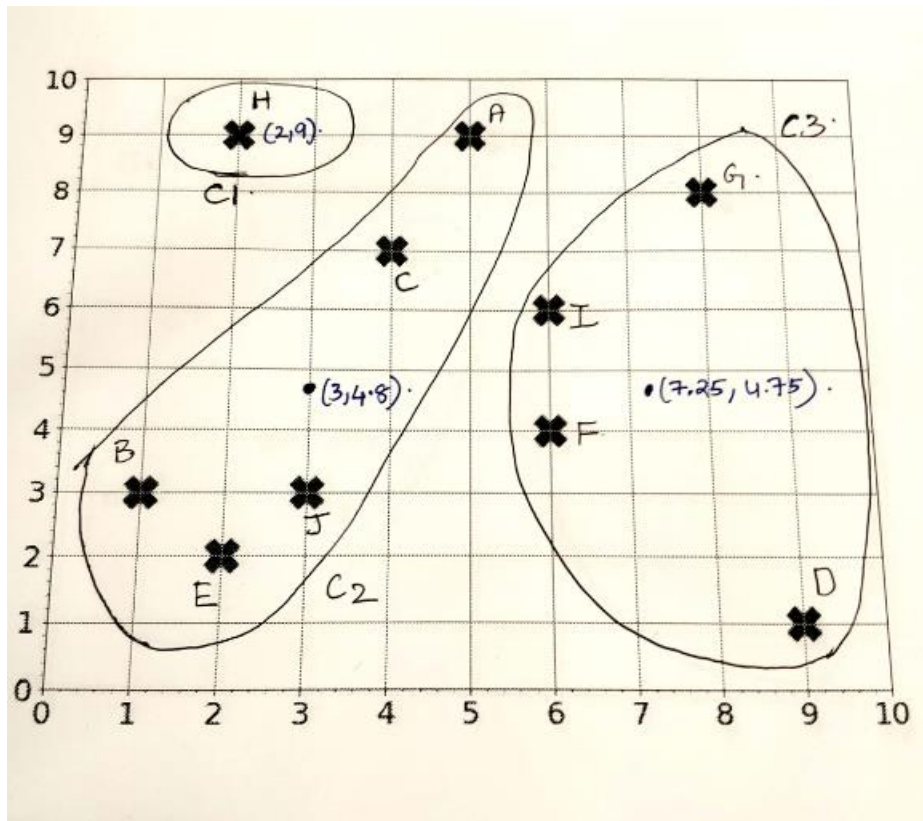


Fig1: Clusters after Round 1 with Centroids shown in blue

## Round-2:

	A	B	C	D	E	F	G	H	I	J
(2,9)	3.00	6.08	2.83	10.63	7.00	6.40	6.08	0	5.00	6.08
(3, 4.8)	4.651	2.69	2.41	7.102	2.97	3.104	5.93	4.317	3.23	1.8
(7.25,4.75)	4.808	6.49	3.952	4.138	5.926	1.457	3.335	6.754	1.767	4.596

### Cluster Members:

C-1 members: {H,A}

C-2 members: { B, C, E, J }

C-3 members: {D,F,G,I}

### New Centroids:

C-1 Centroid: (3.5,9)

C-2 Centroid: (2.5, 3.75)

C-3 Centroid: (7.25,4.75)

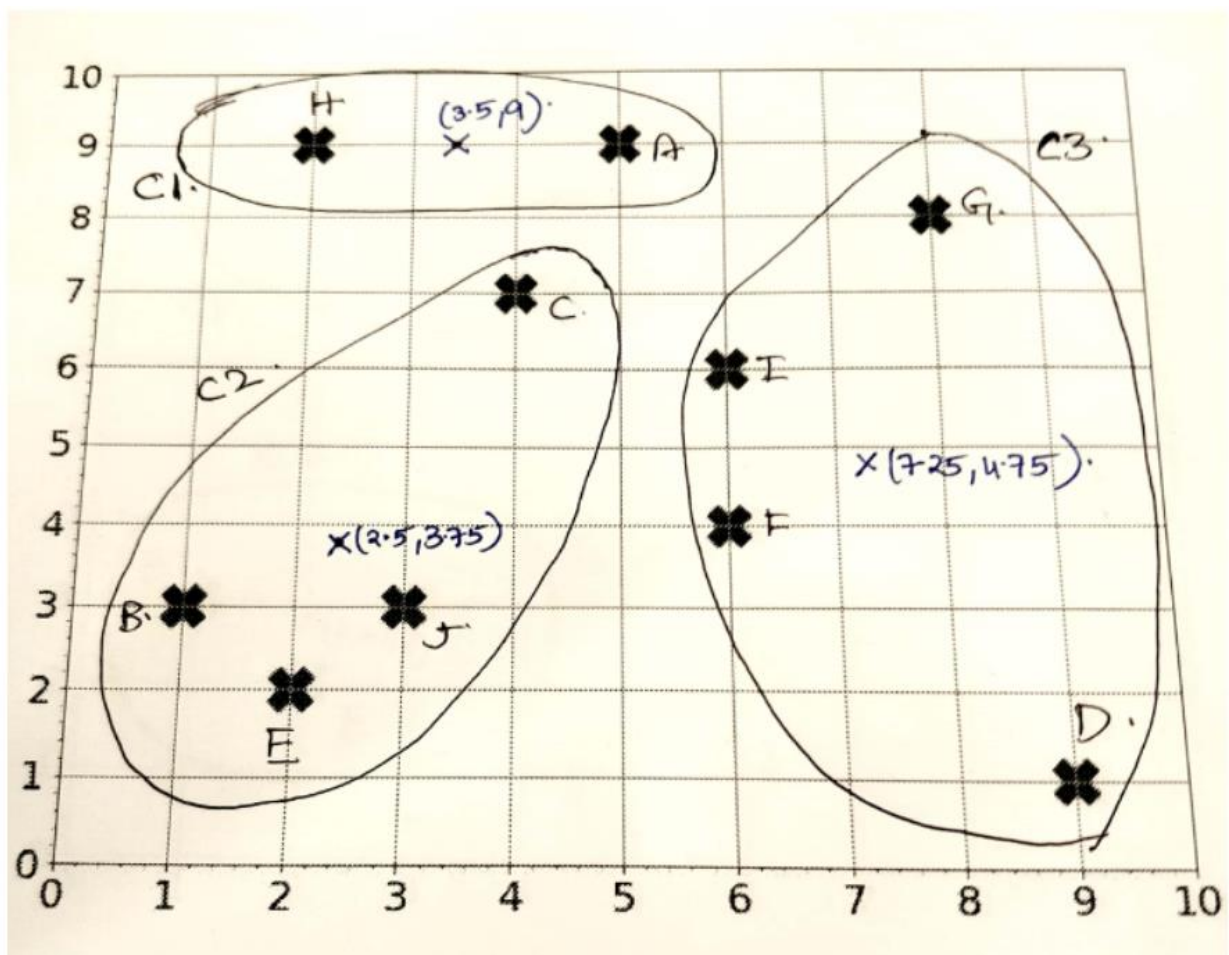


Fig 2: Clusters formed after Round 2 with new Centroids.

### Round 3:

	A	B	C	D	E	F	G	H	I	J
(3.5,9)	1.5	6.5	2.06	9.70	7.15	5.59	4.60	1.5	3.905	6.02
(2.5, 3.75)	5.814	1.677	3.579	7.057	1.820	3.508	6.95	5.273	4.16	0.901
(7.25,4.75)	4.808	6.49	3.952	4.138	5.926	1.457	3.335	6.754	1.767	4.596

#### Cluster Members:

C-1 members: {H,A,C}

C-2 members: { B, E, J}

C-3 members: {D,F,G,I}

#### New Centroids:

C-1 Centroid: (3.67, 8.33)

C-2 Centroid: (2, 2.67)

C-3 Centroid: (7.25,4.75)

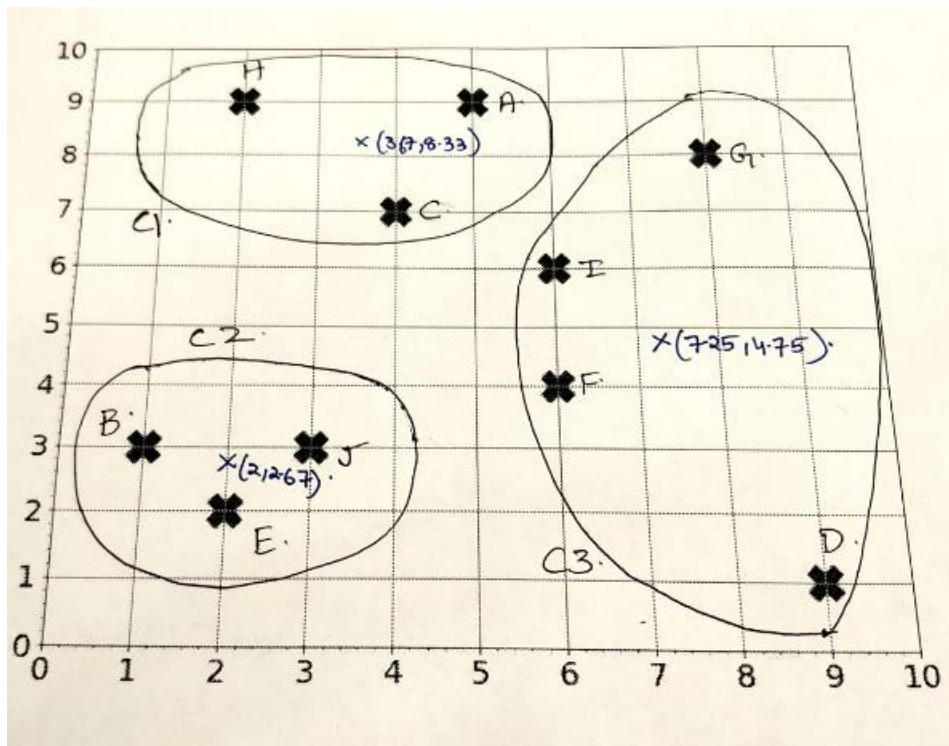


Fig3: Clusters formed after Round 3 with new Centroids.

#### Round 4:

	A	B	C	D	E	F	G	H	I	J
(3.67, 8.33)	1.489	5.96	1.370	9.06	6.54	4.91	4.342	1.799	3.294	5.371
(2, 2.67)	7.004	1.053	4.769	7.196	0.67	4.215	8.025	6.33	5.20	1.05
(7.25, 4.75)	4.808	6.49	3.952	4.138	5.926	1.457	3.335	6.754	1.767	4.596

Cluster Members:

C-1 members: {H,A,C}

C-2 members: { B, E, J}

C-3 members: {D,F,G,I}

We see that the cluster members from Round 4 and Round 3 are same. So the final clusters will be the same as seen in Round 3.

Q1b)

3 rounds were needed for the K-means algorithm to cluster the above data.

Q2)

#### a. Single Link Clustering

	1	2	3	4	5	6	7	8	9	10
1	0	7.21	2.24	8.94	7.62	5.10	3.16	3.00	3.16	6.32
2		0	5.00	8.25	1.41	5.10	8.60	6.08	5.83	2.00
3			0	7.81	5.39	3.61	4.12	2.83	2.24	4.12
4				0	7.07	4.24	7.07	10.63	5.83	6.32
5					0	4.47	8.49	7.00	5.66	1.41
6						0	4.47	6.40	2.00	3.16



Now we calculate the distance of cluster (2,5,10) with other points :

$$\begin{aligned} d((2,5,10),1) &= \min( d((2,5),1),d(10,1)) = \min(7.21,6.32) = 6.32 \\ d((2,5,10),3) &= \min( d((2,5),3),d(10,3)) = \min(5.00,4.12) = 4.12 \\ d((2,5,10),4) &= \min( d((2,5),4),d(10,4)) = \min(6.32,7.07) = 6.32 \\ d((2,5,10),6) &= \min( d((2,5),6),d(10,6)) = \min(3.16,4.47) = 3.16 \\ d((2,5,10),7) &= \min( d((2,5),7),d(10,7)) = \min(7.07,8.49) = 7.07 \\ d((2,5,10),8) &= \min( d((2,5),8),d(10,8)) = \min(6.08,6.08) = 6.08 \\ d((2,5,10),9) &= \min( d((2,5),9),d(10,9)) = \min(4.24,5.66) = 4.24 \end{aligned}$$

	1	(2,5,10)	3	4	6	7	8	9
1	0	6.32	2.24	8.94	5.10	3.16	3.00	3.16
(2,5,10)		0	4.12	6.32	3.16	7.07	6.08	4.24
3			0	7.81	3.61	4.12	2.83	2.24
4				0	4.24	7.07	10.63	5.83
6					0	4.47	6.40	2.00
7						0	6.08	2.83
8							0	5.00
9								0

Now we calculate the distance of cluster (6,9) with other points :

$$\begin{aligned} d((6,9),1) &= \min( d(6,1),d(9,1)) = \min(5.10,3.16) = 3.16 \\ d((6,9),(2,5,10)) &= \min( d(6,(2,5,10)),d(9,(2,5,10))) = \min(4.24,3.16) = 3.16 \\ d((6,9),3) &= \min( d(6,3),d(9,3)) = \min(3.61,2.24) = 2.24 \\ d((6,9),4) &= \min( d(6,4),d(9,4)) = \min(4.24,5.83) = 4.24 \\ d((6,9),7) &= \min( d(6,7),d(9,7)) = \min(4.47,2.83) = 2.83 \\ d((6,9),8) &= \min( d(6,8),d(9,8)) = \min(6.40,5.00) = 5.00 \end{aligned}$$

	1	(2,5,10)	3	4	(6,9)	7	8
1	0	6.32	2.24	8.94	3.16	3.16	3.00
(2,5,10)		0	4.12	6.32	3.16	7.07	6.08
3			0	7.81	2.24	4.12	2.83
4				0	4.24	7.07	10.63

(6,9)					0	2.83	5.00
7						0	6.08
8							0

Now we calculate the distance of cluster (1,3) with other points :

$$d((1,3),(2,5,10)) = \min(6.32,4.12) = 4.12$$

$$d((1,3),4) = \min(d(4,3),d(1,4)) = \min(7.81,8.94) = 7.81$$

$$d((6,9),(1,3)) = \min(3.16,2.24) = 2.24$$

$$d((1,3),7) = \min(d(1,7),d(3,7)) = \min(3.16,4.12) = 3.16$$

$$d((1,3),8) = \min(d(1,8),d(3,8)) = \min(2.83,3.00) = 2.83$$

	(1,3)	(2,5,10)	4	(6,9)	7	8
(1,3)	0	4.12	7.81	2.24	3.16	2.83
(2,5,10)		0	6.32	3.16	7.07	6.08
4			0	4.24	7.07	10.63
(6,9)				0	2.83	5.00
7					0	6.08
8						0

After grouping (1,3) and (6,9), we get (1,3,6,9) . We fill up the proximity matrix :

$$d((1,3,6,9),(2,5,10)) = \min(4.12,3.16) = 3.16$$

$$d((1,3,6,9),4) = \min(7.81,4.24) = 4.24$$

$$d((1,3,6,9),7) = \min(3.16,2.83) = 2.83$$

$$d((1,3,6,9),8) = \min(2.83,5.00) = 2.83$$

	(1,3,6,9)	(2,5,10)	4	7	8
(1,3,6,9)	0	3.16	4.24	2.83	2.83
(2,5,10)		0	6.32	7.07	6.08
4			0	7.07	10.63
7				0	6.08
8					0

Now we calculate the distance of cluster (1,3,6,9,7) :



$$d((1,3,6,9,7),(2,5,10)) = \min(3.16, 7.07) = 3.16$$

$$d((1,3,6,9,7),4) = \min(4.24, 7.07) = 4.24$$

$$d((1,3,6,9,7),8) = \min(2.83, 6.08) = 2.83$$

	(1,3,6,9,7)	(2,5,10)	4	8
(1,3,6,9,7)	0	3.16	4.24	2.83
(2,5,10)		0	6.32	6.08
4			0	10.63
8				0

Next group is (1,3,6,9,7,8):

$$d((1,3,6,9,7,8),(2,5,10)) = \min(3.16, 6.08) = 3.16$$

$$d((1,3,6,9,7,8),4) = \min(4.24, 10.63) = 4.24$$

	(1,3,6,9,7,8)	(2,5,10)	4
(1,3,6,9,7,8)	0	3.16	4.24
(2,5,10)		0	6.32
4			0

$$d((1,3,6,9,7,8,2,5,10),4) = \min(4.24, 6.32) = 4.24$$

	(1,3,6,9,7,8,2,5,10)	4
(1,3,6,9,7,8,2,5,10)	0	4.24
4		0

Finally we merge (1,3,6,9,7,8,2,5,10) with 4 at distance of 4.24.

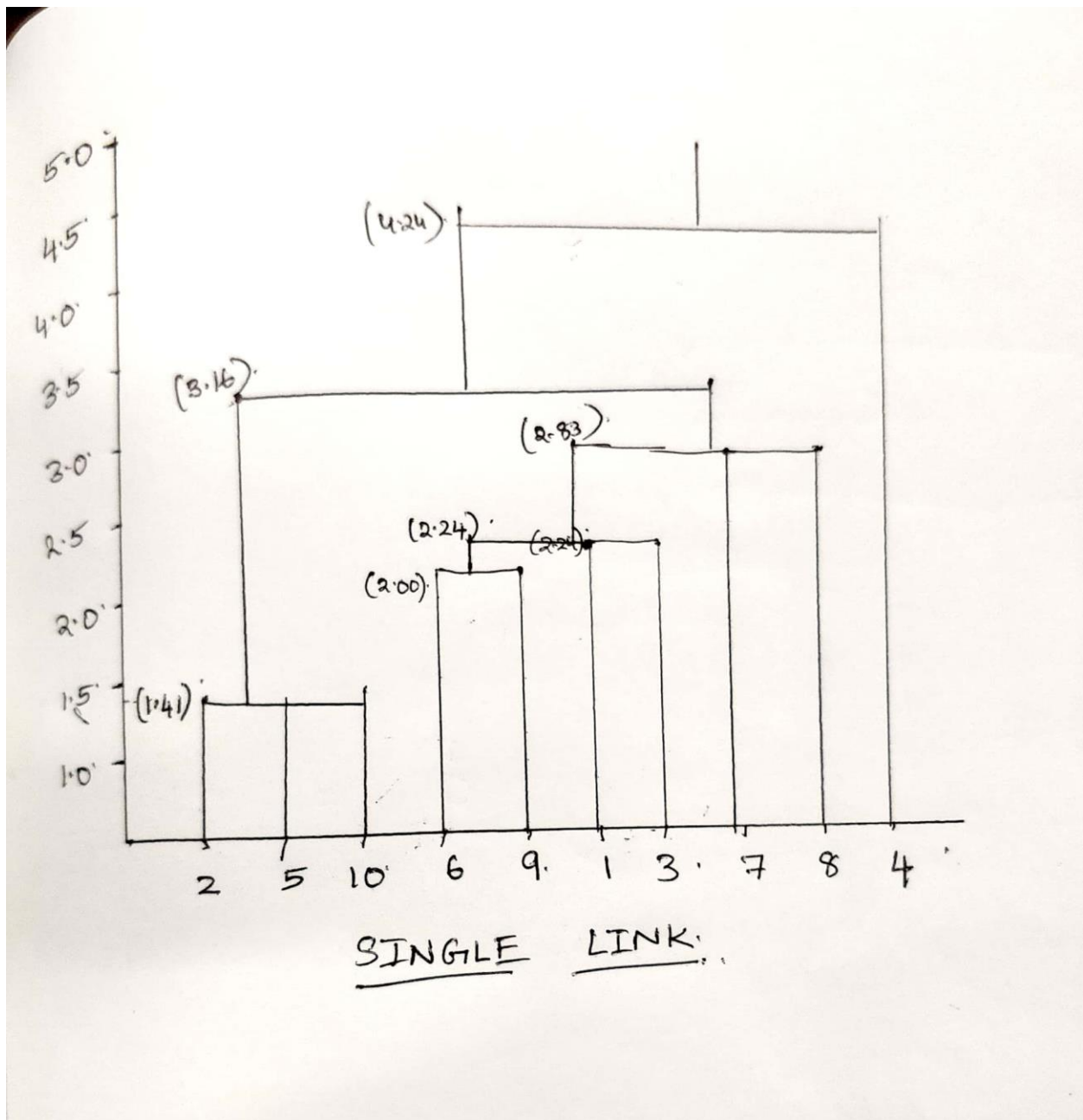


Fig 4: Dendrogram for Single Link Hierarchical Clustering

## Q2 b) Complete Link Clustering

	1	2	3	4	5	6	7	8	9	10
1	0	7.21	2.24	8.94	7.62	5.10	3.16	3.00	3.16	6.32
2		0	5.00	8.25	1.41	5.10	8.60	6.08	5.83	2.00
3			0	7.81	5.39	3.61	4.12	2.83	2.24	4.12
4				0	7.07	4.24	7.07	10.63	5.83	6.32
5					0	4.47	8.49	7.00	5.66	1.41
6						0	4.47	6.40	2.00	3.16
7							0	6.08	2.83	7.07
8								0	5.00	6.08
9									0	4.24
10										0

Now we calculate the distance of cluster (2,5) using the Complete link hierarchical clustering.

$$\begin{aligned}
 d((2,5),1) &= \max(d(2,1), d(5,1)) = \max(7.21, 7.62) = 7.62 \\
 d((2,5),3) &= \max(d(2,3), d(5,3)) = \max(5.00, 5.39) = 5.39 \\
 d((2,5),4) &= \max(d(2,4), d(5,4)) = \max(8.25, 7.07) = 8.25 \\
 d((2,5),6) &= \max(d(2,6), d(5,6)) = \max(5.10, 4.47) = 5.10 \\
 d((2,5),7) &= \max(d(2,7), d(5,7)) = \max(8.60, 8.49) = 8.60 \\
 d((2,5),8) &= \max(d(2,8), d(5,8)) = \max(6.08, 7.00) = 7.00 \\
 d((2,5),9) &= \max(d(2,9), d(5,9)) = \max(5.83, 5.66) = 5.83 \\
 d((2,5),10) &= \min(d(2,10), d(5,10)) = \min(2.00, 1.41) = 2.00
 \end{aligned}$$

	1	(2,5)	3	4	6	7	8	9	10
1	0	7.62	2.24	8.94	5.10	3.16	3.00	3.16	6.32
(2,5)		0	5.39	8.25	5.10	8.60	7.00	5.83	2.00
3			0	7.81	3.61	4.12	2.83	2.24	4.12
4				0	4.24	7.07	10.63	5.83	6.32
6					0	4.47	6.40	2.00	3.16
7						0	6.08	2.83	7.07
8							0	5.00	6.08

9								0	4.24
10									0

$d((2,5,10),1) = \max(d((2,5),1),d(10,1)) = \max(7.62,6.32) = 7.62$   
 $d((2,5,10),3) = \max(d((2,5),3),d(10,3)) = \max(5.39,4.12) = 5.39$   
 $d((2,5,10),4) = \max(d((2,5),4),d(10,4)) = \max(6.32,8.25) = 8.25$   
 $d((2,5,10),6) = \max(d((2,5),6),d(10,6)) = \max(3.16,5.10) = 5.10$   
 $d((2,5,10),7) = \max(d((2,5),7),d(10,7)) = \max(7.07,8.60) = 8.60$   
 $d((2,5,10),8) = \max(d((2,5),8),d(10,8)) = \max(6.08,7.00) = 7.00$   
 $d((2,5,10),9) = \max(d((2,5),9),d(10,9)) = \max(4.24,5.83) = 5.83$

	1	(2,5,10)	3	4	6	7	8	9
1	0	7.62	2.24	8.94	5.10	3.16	3.00	3.16
(2,5,10)		0	5.39	8.25	5.10	8.60	7.00	5.83
3			0	7.81	3.61	4.12	2.83	2.24
4				0	4.24	7.07	10.63	5.83
6					0	4.47	6.40	2.00
7						0	6.08	2.83
8							0	5.00
9								0

Now we calculate the distance of cluster (6,9) with other points :

$d((6,9),1) = \max(d(6,1),d(9,1)) = \max(5.10,3.16) = 5.10$   
 $d((6,9),(2,5,10)) = \max(d(6,(2,5,10)),d(9,(2,5,10))) = \max(5.10,5.83) = 5.83$   
 $d((6,9),3) = \max(d(6,3),d(9,3)) = \max(3.61,2.24) = 3.61$   
 $d((6,9),4) = \max(d(6,4),d(9,4)) = \max(4.24,5.83) = 5.83$   
 $d((6,9),7) = \max(d(6,7),d(9,7)) = \max(4.47,2.83) = 4.47$   
 $d((6,9),8) = \max(d(6,8),d(9,8)) = \max(6.40,5.00) = 6.40$

	1	(2,5,10)	3	4	(6,9)	7	8
1	0	7.62	2.24	8.94	5.10	3.16	3.00
(2,5,10)		0	5.39	8.25	5.83	8.60	7.00
3			0	7.81	3.61	4.12	2.83

4				0	5.83	7.07	10.63
(6,9)					0	4.47	6.40
7						0	6.08
8							0

Now we calculate the distance of cluster (1,3) with other points :

$$d((1,3),(2,5,10)) = \max(7.62, 5.39) = 7.62$$

$$d((1,3),4) = \max(d(4,3), d(1,4)) = \max(7.81, 8.94) = 8.94$$

$$d((6,9),(1,3)) = \max(5.10, 3.61) = 5.10$$

$$d((1,3),7) = \max(d(1,7), d(3,7)) = \max(3.16, 4.12) = 4.12$$

$$d((1,3),8) = \max(d(1,8), d(3,8)) = \max(2.83, 3.00) = 3.00$$

	(1,3)	(2,5,10)	4	(6,9)	7	8
(1,3)	0	7.62	8.94	5.10	4.12	3.00
(2,5,10)		0	8.25	5.83	8.60	7.00
4			0	5.83	7.07	10.63
(6,9)				0	4.47	6.40
7					0	6.08
8						0

Now we calculate the distance of cluster (1,3,8) with other points/clusters :

$$d((1,3,8),(2,5,10)) = \max(7.62, 7.00) = 7.62$$

$$d((1,3,8),4) = \max(7.81, 10.63) = 10.63$$

$$d((6,9),(1,3,8)) = \max(5.10, 6.40) = 6.40$$

$$d((1,3,8),7) = \max(6.08, 4.12) = 6.08$$

	(1,3,8)	(2,5,10)	4	(6,9)	7
(1,3,8)	0	7.62	10.63	6.40	6.08
(2,5,10)		0	8.25	5.83	8.60
4			0	5.83	7.07
(6,9)				0	4.47
7					0

Now we calculate the distance of cluster (6,9,7) with other points/clusters :

$$d((6,9,7),(2,5,10)) = \max(5.83,8.60) = 8.60$$

$$d((6,9,7),4) = \max(5.83,7.07) = 7.07$$

$$d((6,9,7),(1,3,8)) = \max(6.08,6.40) = 6.40$$

	(1,3,8)	(2,5,10)	4	(6,9,7)
(1,3,8)	0	7.62	10.63	6.40
(2,5,10)		0	8.25	8.60
4			0	7.07
(6,9,7)				0

$$d((1,3,8,6,9,7), (2,5,10)) = \max(7.62,8.60) = 8.60$$

$$d((1,3,8,6,9,7),4) = \max(10.63,7.07) = 10.63$$

	(1,3,8,6,9,7)	(2,5,10)	4
(1,3,8,6,9,7)	0	8.60	10.63
(2,5,10)		0	8.25
4			0

$$d((2,5,10,4),(1,3,8,6,9,7)) = \max(8.60,10.63) = 10.63$$

	(1,3,8,6,9,7)	(2,5,10,4)
(1,3,8,6,9,7)	0	10.63
(2,5,10,4)		0

Finally (1,3,8,6,9,7) will be clustered together along with (2,5,10,4) at distance of 10.63.

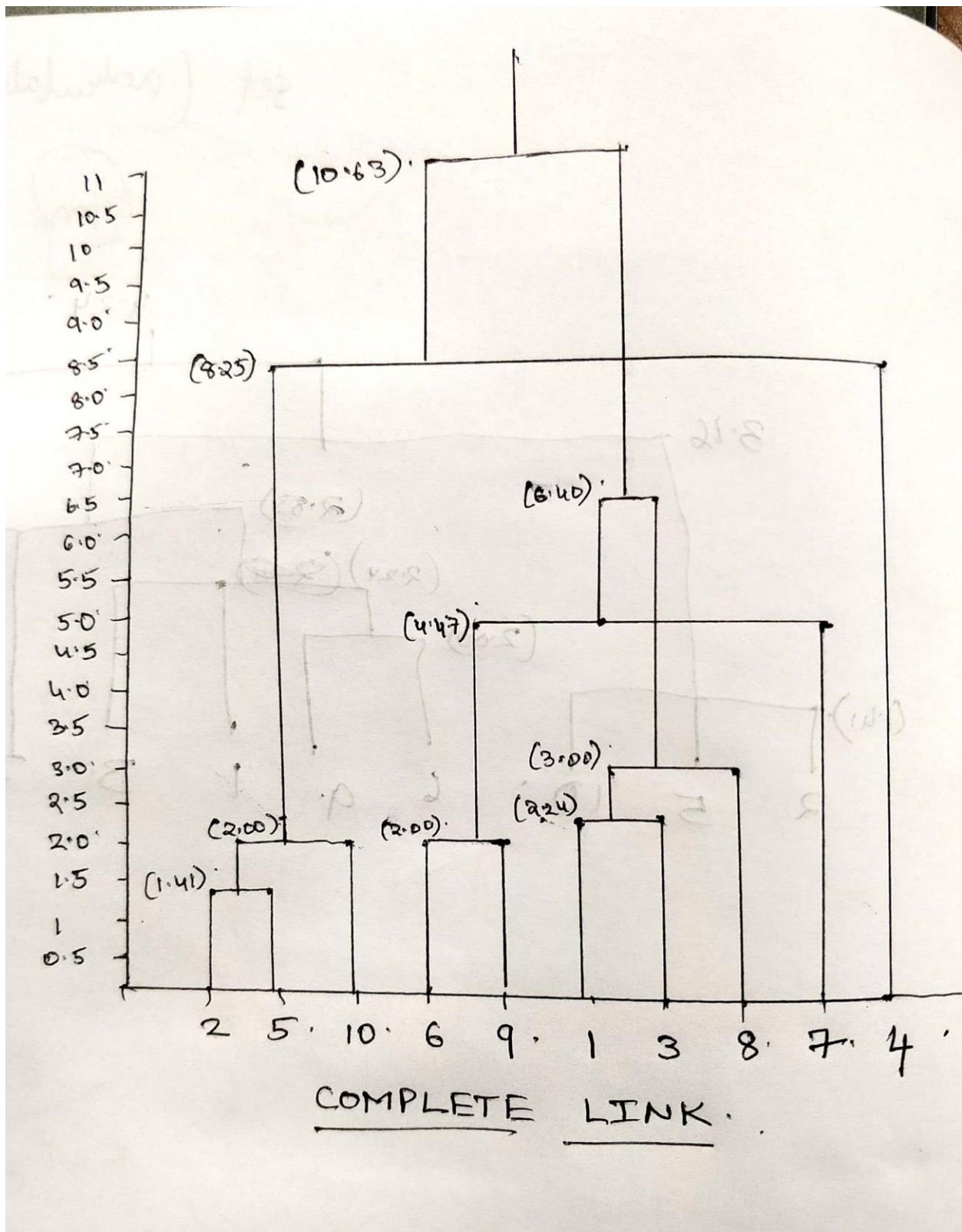


Fig 5: Dendrogram for Complete Link Hierarchical Clustering

Q2)

c) The following are the properties of the Single link and Complete link clusters:

Single link cluster:

- It can handle non-elliptical cluster shapes.
- Single-link clustering is not favourable when there is any noise or outliers in the data.
- It doesn't give a good balanced structure for the nested clusters compared to complete link.

Complete link cluster:

- On the contrary of Single link cluster, complete link clustering gives balanced structure for the nested clusters.
- It is less susceptible to noise and outliers since we are considering farthest distance.
- Doesn't handle the large or non-elliptical clusters properly.

So, when there are two clusters and if they contain outliers or noise complete link hierarchical clustering approach gives better clustering. If the clusters are large or non-elliptical shapes than single link hierarchical clustering approach is better.

#### **d) Interpretation of Silhouette Coefficient:**

A higher value of the Silhouette coefficient indicates that the data point is far away from the neighbouring clusters. A value of 0 indicates that the data point is very close to the decision boundary between two neighbouring clusters. A negative value is an indication that the data point has been assigned to the wrong cluster.

The Silhouette coefficient for the Single Link Hierarchical Clustering is obtained to be 0.39 and the Silhouette coefficient for the K-means clustering is obtained to be 0.417. Higher value of the Silhouette Coefficient for Single Link Hierarchical Clustering indicates that it is better clustered. Yes, we agree with this assessment. A visual inspection over the 2-D plot of the points indicates that the clusters formed by Single Link Hierarchical Clustering are better grouped together.

Below is the Code used for obtaining the Silhouette Coefficient:



```
from sklearn.metrics import silhouette_score
from sklearn.metrics.pairwise import euclidean_distances
```

```
X = [[5, 9], [1, 3],[4, 7], [9, 1],[2, 2], [6, 4],[8, 8], [2, 9],[6, 6], [3, 3]]
XX = euclidean_distances(X, X)
```

```
# for single link hierarchical clustering
Y = [0,1,0,2,1,2,2,0,2,1]
sil = silhouette_score(XX, Y, metric='precomputed')
print(sil)
```

```
0.39033931899857366
```

```
# for k-means clustering
YY = [1,3,1,2,3,1,1,1,3]
sil2 = silhouette_score(XX, YY, metric='precomputed')
print(sil2)
```

```
0.41727528475909487
```

Q3)

a) Maximum number of unique itemsets that can be extracted = 36

1-itemsets: 6

{Bread}, {Milk}, {Butter}, {Eggs}, {Beer}, {Cola}

2-itemsets: 14

{Bread, Milk}, {Bread, Cola}, {Milk, Cola}, {Bread, Butter}, {Bread, Eggs},  
{Bread, Beer}, {Butter, Eggs}, {Butter, Beer}, {Eggs, Beer}, {Milk, Butter}, {Milk, Eggs},  
{Butter, Cola}, {Beer, Cola}, {Eggs, Cola}

3-itemsets: 13

{Bread, Milk, Cola}, {Bread, Butter, Eggs}, {Bread, Butter, Beer}, {Bread, Eggs, Beer},  
{Milk, Butter, Eggs}, {Milk, Butter, Cola}, {Butter, Eggs, Cola}, {Milk, Eggs, Cola}, {Bread, Milk,  
Eggs}, {Bread, Milk, Butter}, {Bread, Butter, Cola}, {Butter, Beer, Cola}

4-itemsets: 3

{Bread, Butter, Eggs, Beer}

{Milk, Butter, Eggs, Cola}

{Bread, Milk, Butter, Cola}

b) Maximum number of association rules possible on the given dataset is given by:

$$3^d - 2^{d+1} + 1 = 3^6 - 2^{(6+1)} + 1 = 602 \text{ (d = 6)}$$

c) Support count of {Eggs,Cola} in the given dataset is 2. Therefore the support of the itemset {Eggs,Cola} is:

$$\sigma(\{Eggs,Cola\}) / |T| = 2/10 = 0.2$$

d) Support for the association rule {Bread} -> {Butter} is given by:

$$\sigma(\{\text{Bread, Butter}\}) / |T| = 3/10 = 0.3$$

Confidence of the association rule {Bread} -> {Butter} is given by:

$$\sigma(\{\text{Bread, Butter}\}) / \sigma(\{\text{Bread}\}) = 3/6 = 0.5$$

e) All valid association rules of the form {A, B} -> {C} having min support = 0.3 and min confidence = 0.6 are:

{Bread, Milk} -> {Cola}

{Bread, Cola} -> {Milk}

{Milk, Cola} -> {Bread}

f) With surety, we can confirm that the support of the association rule {a} -> {b,d} is between 0.23 and 0.46 because assuming there are 100 itemsets and it is given that {a,b} is present in 46 of them and {a,b,c,d} is present in 23 of them, then {a,b,d} will be present in [23, 46]. Hence the support of {a} -> {b,d} lies between 0.23 and 0.46.

Q4)

a)

Dataset

TID	ITEMS
t1	A,B,C,D
t2	A,B,D,E
t3	A,B
t4	A,C,D
t5	A,C,E
t6	B,C
t7	C,D
t8	C,D,E

C1

Itemset	Sup
{A}	5
{B}	4
{C}	6
{D}	5
{E}	3

L1

Itemset	Sup
{A}	5
{B}	4
{C}	6
{D}	5
{E}	3

C2

Itemset	Sup
{A,B}	3
{A,C}	3
{A,D}	3
{A,E}	2
{B,C}	2
{B,D}	2
{B,E}	1

{C,D}	4
{C,E}	2
{D,E}	2

L2

Itemset	Sup
{A,B}	3
{A,C}	3
{A,D}	3
{C,D}	4

C3

Itemset	Sup
{A,C,D}	2

We have the initial dataset D. After the first scan C1 , we calculate the frequent 1-itemsets which have min support of 3. In this scan, we observe all itemsets have minimum threshold, therefore all the 1-itemsets [ {A},{B},{C},{D},{E} ] are frequent. This can be seen in L1.

In the next step we scan L1 in the step C2 where we try to form frequent 2-itemsets. Here we observe that the itemsets [ { A,B}, {A,C},{A,D},{C,D} ] are the only itemsets which have a min-support count of 3. So only these itemsets are considered and taken in L2, the rest are ignored.

A scan on L2 , i.e C3 is used to find the frequent 3 itemsets. Here there can be only 1 valid itemset that can be joined and formed from L2. That itemset is {A,C,D}. But the min-support count of the itemset {A,C,D} is found to be 2 which is below our required threshold. Hence there are no further itemsets to be considered and we stop our Apriori process here.

The frequent itemsets are [ {A}, {B}, {C}, {D}, {E}, {A,B}, {A,C}, {A,D}, {C,D} ].

Q4)

b) Lattice diagram

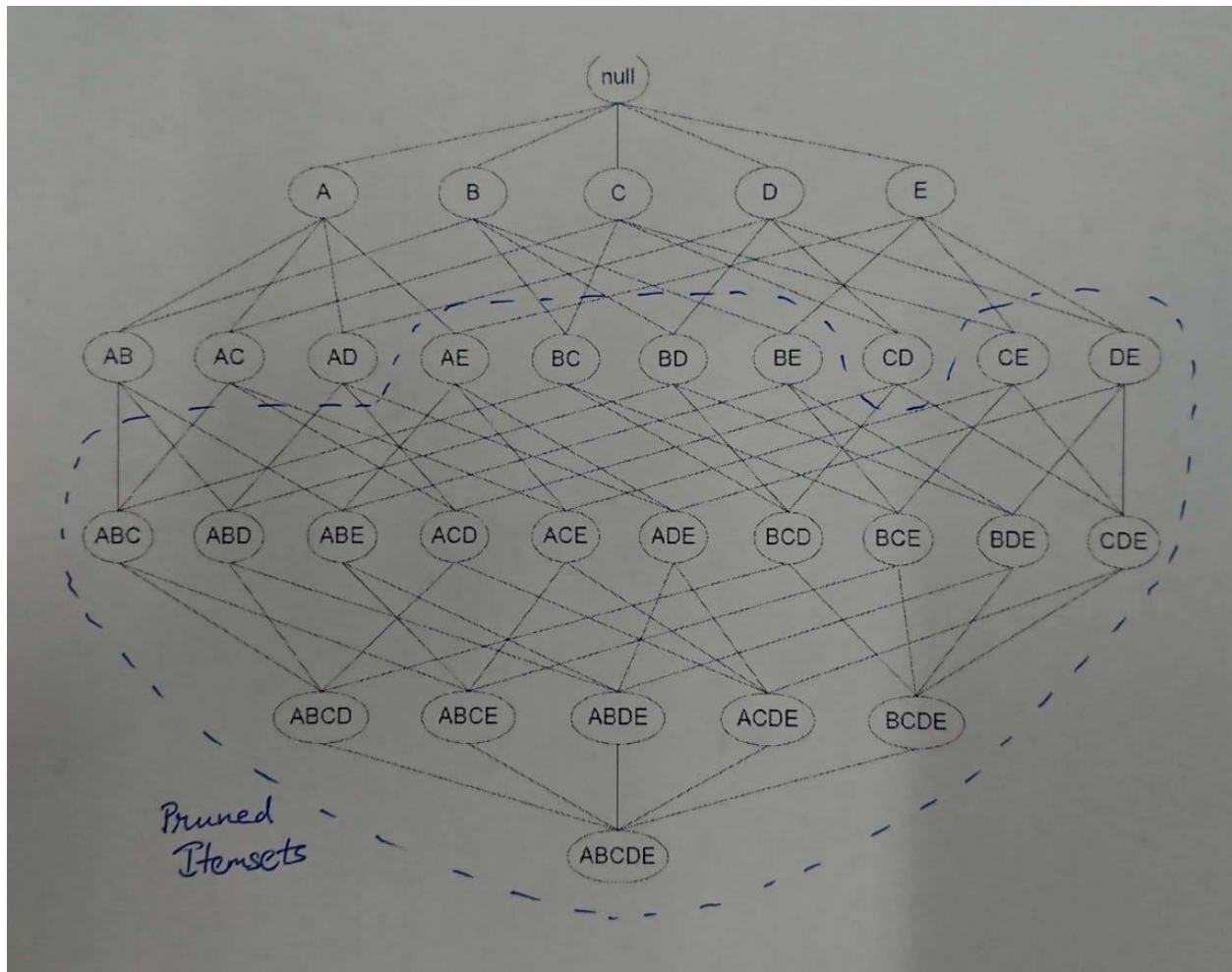
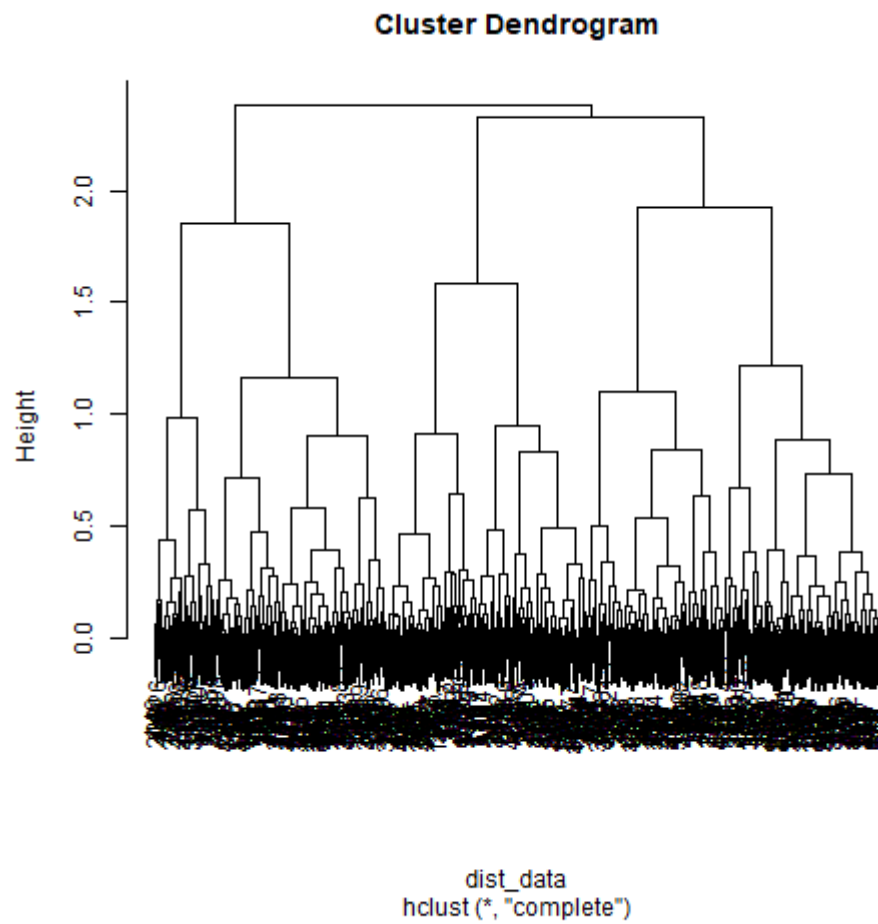


Fig 6: Dotted lines show the boundary for pruned itemsets.

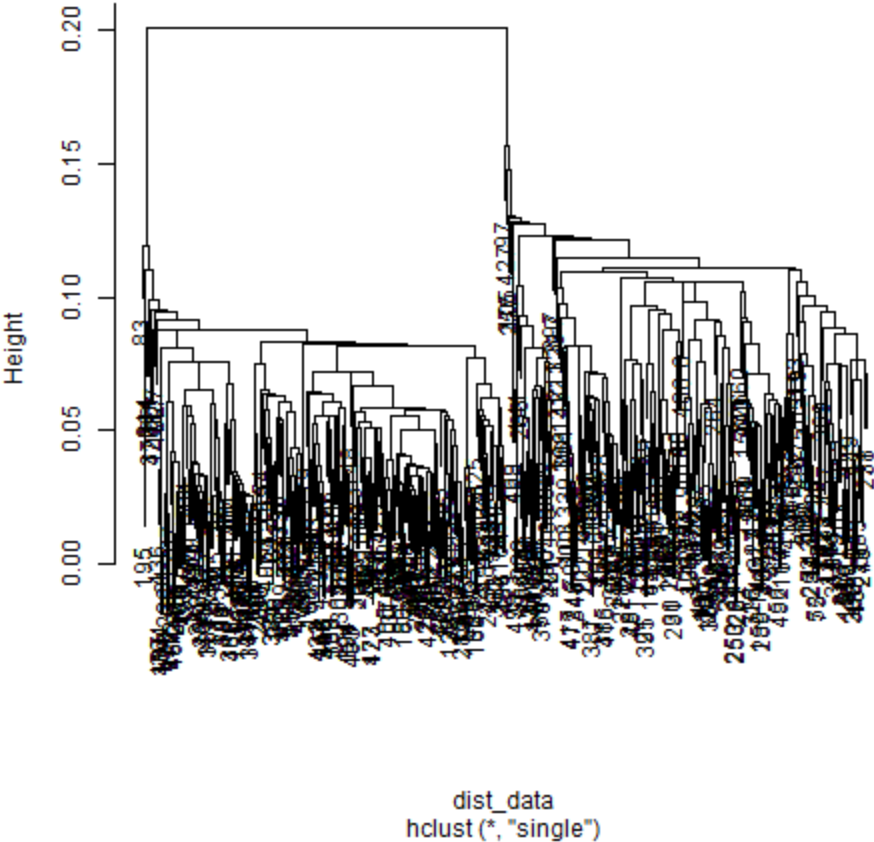
In the above lattice, the item sets below the dotted line are the pruned itemsets and non-frequent itemsets.

Q5)

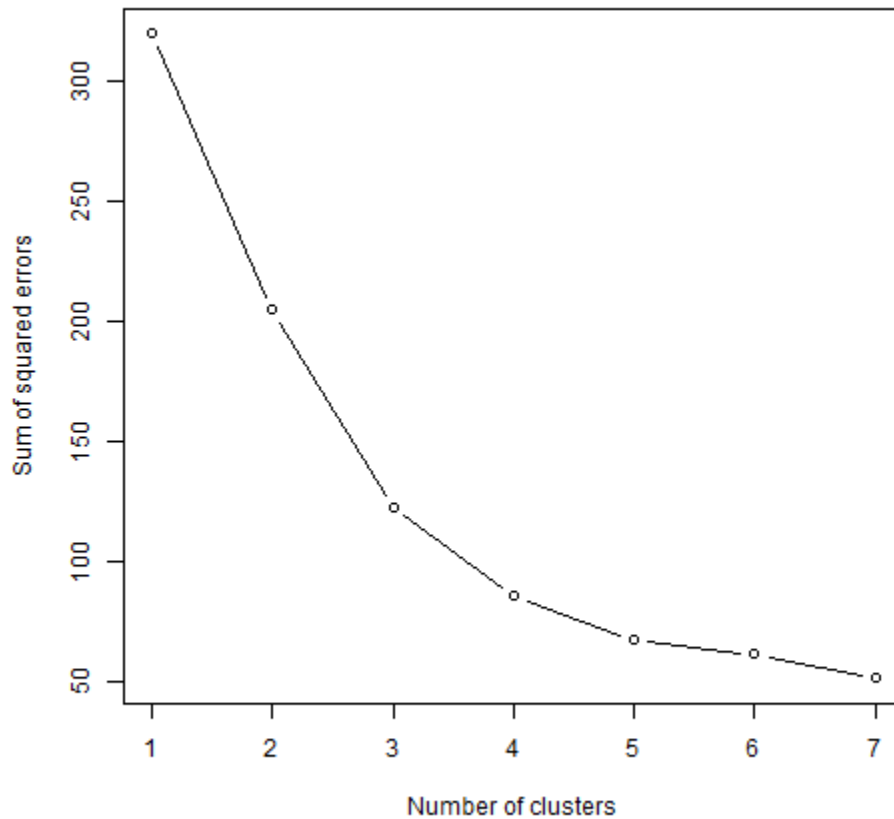
A. Dendrogram for complete and single link clustering :



Cluster Dendrogram



C) The elbow plot generated is as given below :



Looking at this elbow plot where the SSE is relatively small, but not too small so as to overfit. As per observation, this point is at  $k = 4$ . As we see that the SSE has considerably decreased when we go from  $k = 3$  to  $k = 4$ , plus there is not much decrease in SSE when we go from  $k = 4$  to  $k = 5$ . Hence the best value of  $k$  would be 4.

D) The SSE values from the code are below :

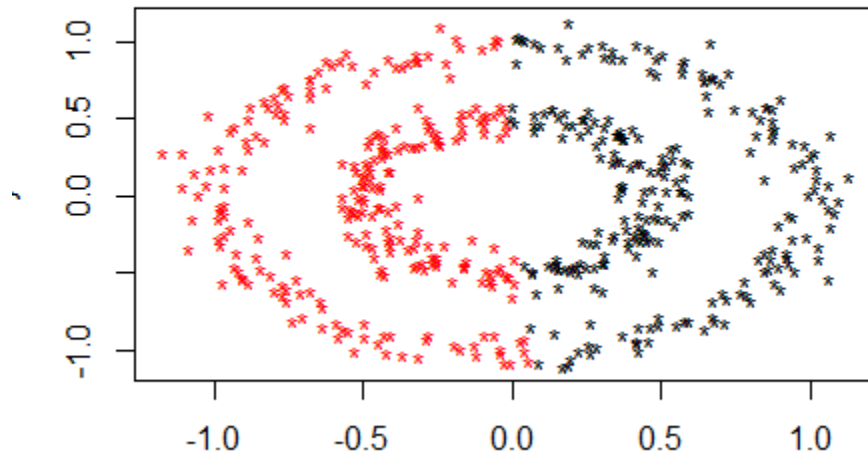
```
[1] "kmeans SSE for given params = 204.626515194937"  
> print(paste("single link SSE for given params = ", single_link_sse))  
[1] "single link SSE for given params = 319.693456992432"  
> print(paste("complete link SSE for given params = ", complete_link_sse))  
[1] "complete link SSE for given params = 221.04197449858"
```

Purely based on SSE, it is intuitive to go with the algorithm having least SSE. Hence based purely on SSE, we would choose Kmeans.

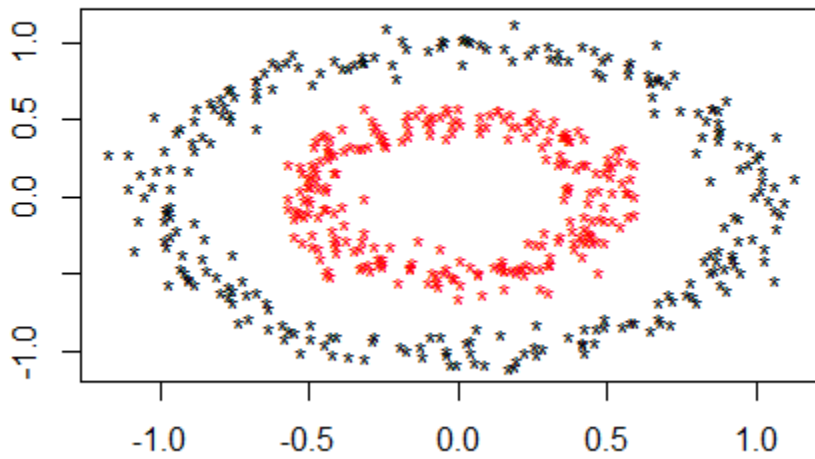


E) The 3 cluster plots are given below :

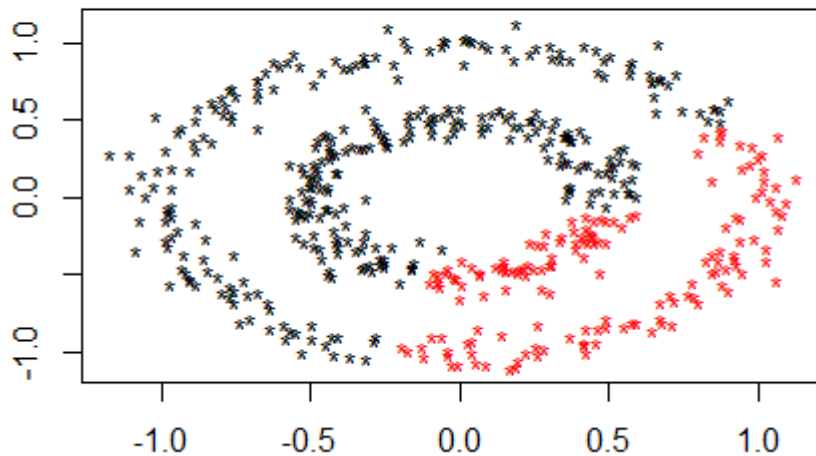
**KMeans with 2 clusters**



**Single Link with 2 clusters**



### Complete link with 2 clusters



Based on the plot, and the shape of the data points, the Single Link clustering method gives good and visually intuitive separation of the data points as shown in the plot. The other two clustering methods do not seem to have a clear separation between the data points which are grouped into their respective clusters. Hence, via the figures, single link clustering is the best method.

F) No, the answers for the visual comparison and the SSE comparison differ. The conclusions we can draw from this is the underlying data might have a possible hierarchical structure as single-link gives better visualization than k means for the same. Kmeans does not react well to unusual shaped clusters. Visualizations give a more intuitive cluster assignment if the data is two dimensional. This might change for more dimensional data, and then in that case we will have to rely more on numeric measures to judge cluster quality. Numeric measures might not always be reliable. It might be the case that the cluster assignments overfit to the data, hence giving us a low SSE count.

(b)

iii)

A. The decay and number of units in the hidden layer is as given below :

Accuracy was used to select the optimal model using the largest value.  
The final values used for the model were size = 16 and decay = 0.5.